Cti AI llm training

Generative AI OSINT and Cyber Intelligence.

The Observable Surface web has so far not yet been comprehensively mapped mathematically by Multimodal large language/learning models. There was an interesting observation by Janes that around 2 decades back the quantum of Open source Intelligence was just about the size of the salt one may sprinkle on its meals. Today it represents about 95% of all cyber intel gathering and analysis driven research . We may also add the AI enabled Algorithmic analysis to that trend for better , valuable and vivid insights .

The deficiencies of generative Intelligence range from the impairment observed by virtue of the human overseers maybe on account of enforcing guardrails . A recent example that may be cited is the prompt to know what percentage of the global annual defense budget of countries is allocated towards the development and improvisation of state of art , versatile, adaptable , both centralised and decentralised-distrubuted edge computing powered Autonomous Weapons systems and platforms (LAWS - UAVs UGVs) etc. Also the relevant Government and Military policies towards this..the approach atleast...The prompt so far couldn't fetch straight forward answers or succinct analysis based on facts from the surface web atleast.

Here Machine based generative intelligence marked the sensitive nature of the above defense technology...But on the contrary plenitude is available on the surface web through absolutely official and legal channels .

Second instance that can be cited is the query on Advanced integrated modular portable battlefield management systems cum ecosystems and the Military metaverse. Again Whole lot of direct information is available in form of detailed reports on the surface web through absolutely official and non classified channels.

So we believe the human cyber intelligence and defense analysts will need to pre process this information in tabular arrays or grids for training multi modal generative AI LLMs. Also mapping the current cyber threat landscape through navigation of the web portals , reports , documents.

In order of importance for learning through data ingestion AI LLMs prefer not dense but significantly plain simple English content (that day I was reading the Zombie Noun Apocalypse in Military Defense Doctrine writing) . So to begin with..Well organised and paraphrased linear textual notes...next it prefers preprocessed information neatly organised in tabular arrays or grids .

As Spreadsheets and tables formats power the World's most important both legacy and Cloud based software as a service ecosystem - The Relational Database management systems and Ms office Excel and free and open source Libreoffice..Google sheets of Google docs etc.

Establishing associations and links between keywords , key concepts for pattern recognition remains a human analysts job . Just like we do for our mind maps.

Self aware Autonomous cyber defence systems should be in a position to conduct OSINT ops and gain credible insights into both the surface and dark web to defeat APTs and malicious actors .

Simplifying the entire complex process of training cyber intelligence and OSINT specific AI models can theoretically begin right on our smart devices.

Let's say if one is on the Android ecosystem , and focusing on Google Chrome while safe browsing under enhanced protection which offers real time proactive cyber defense against threats... the data gets teleported to servers of Google . This precisely can be used to build a let's say cyber threat intelligence model especially for cyber analysts who choose or opt to do so.

Not confusing this with the privacy concerns so on and so forth.

So a lot of cyber intelligence training for multimodal generative AI LLMs can take place at the elementary browsing level.

Google Deepmind through project AndroidEnv learns from human operators sessions on Android gaming , navigation maps , calculations , Flight booking and reservations etc. Also at the Android /Linux / Chrome OS granular os level there are a whole lot of settings that can be toggled on for a lot of machine learning to take place through sharing of intelligible data , analytics and insights.

Even Windows and Mac OS /iOS can benefit from this learning pattern for cyber intelligence analysts who opt to train specific LLMs right from their smart devices augmenting the power of cloud computing. We also understand that tremendous amount of learning happens at the granular elementary operating systems intelligence level (machine device learning for relevant people , app , content suggestion)..from app usage times..usage patterns..screen time.. personalisation of keyboard usage statistics..for corrections and accuracy improvement..audio snippets donation..Also the Original equipment manufacturers and OS non encrypted Cloud backup services..user experience and system stability improvement programs..system logs , error analysis , bug reports.. anonymous analysation of general statistical data for improvisation and future development of product design..Device ads id for ad performance measurement etc. Computer device performance and diagnostics are directly proportional to improve battery life and system / app stability..Data aggregation is a boon for AI enabled app development.. Intelligible Data ingestion through decentralised sources does have single point of convergence at data strong Cloud servers.. Therefore theoretically crafting/designing individual AI learning models custom built and fine tuned over specific preferences is a distant possibility over the cyber electronic horizon.

This will eventually trickle down to offline on device AI and billions of parameters just for those who do not want any real time human overseeing and skewing of generative AI's results.

Machines are already augmenting the capability of onboard / ondevice edge AI through powerful combination of System on chip - CPU , GPU and NPU - especially the new gaming series laptops powered by Nvidia , ARMs Snapdragon, INTELs , Samsung Exonys chips.

Securing open source OS and software will also remain a major challenge...Military Intelligence grade OS distros/variants and software components derivatives must achieve immutability.

Finally the last scope for improvement is in the domain of cyber and science based strategic communications which I still believe humans are better at.

(Let's presume that multimodal generative AI starts learning this trait by analysing Zombie Nouns and Dense Military Cyber Intelligence defense penning).

Also at some point AI powered computer vision will have to see by virtue of ocular augmentation the elementary world of subatomic particles , dark energy and dark matter..establish inroads into the field of noetic philosophers and SETI scientists cum explorers .

All Science and Mathematics ultimately begun through Philosophical Abstractionist ideals and principles.

This mammoth Super Quantum powered AI multi modal LLMs will also have to connect human precognition /Clairvoyance , psychokinetics / telekinesis etc to deeply understand Human Encephalon based Cognitive abilities and Consciousness. Computational Neuroscience Neuro informatics and Brain Simulation. For neural simulation the fastest Super Quantum Computing power will have to be mapped on logarithmic scale and exponential treadline.

One possible approach to achieving AGI is whole brain emulation: A brain model is built by scanning and mapping a biological brain in detail, and then copying and simulating it on a computer system or another computational device. The simulation model must be sufficiently faithful to the original, so that it behaves in practically the same way as the original brain. Whole brain emulation is discussed in computational neuroscience and neuroinformatics, in the context of brain simulation for medical research purposes. It has been discussed in artificial intelligence research as an approach to strong AI. Neuroimaging technologies that could deliver the necessary detailed understanding are improving rapidly, and futurist Ray Kurzweil in the book The Singularity Is Near predicts that a map of sufficient quality will become available on a similar timescale to the computing power required to emulate it.

For low-level brain simulation, a very powerful cluster of computers or GPUs would be required, given the enormous quantity of synapses within the human brain. Each of the 1011 (one hundred billion) neurons has

on average 7,000 synaptic connections (synapses) to other neurons. The brain of a three-year-old child has about 1015 synapses (1 quadrillion). This number declines with age, stabilizing by adulthood. Estimates vary for an adult, ranging from 1014 to 5×1014 synapses (100 to 500 trillion). An estimate of the brain's processing power, based on a simple switch model for neuron activity, is around 1014 (100 trillion) synaptic updates per second (SUPS).

In 1997, Kurzweil looked at various estimates for the hardware required to equal the human brain and adopted a figure of 1016 computations per second (cps). (For comparison, if a "computation" was equivalent to one "floating-point operation" – a measure used to rate current supercomputers – then 1016 "computations" would be equivalent to 10 petaFLOPS, achieved in 2011, while 1018 was achieved in 2022.) He used this figure to predict the necessary hardware would be available sometime between 2015 and 2025, if the exponential growth in computer power at the time of writing continued

Only then we can see the emergence of Self Aware AGI and Singularity of Sentinence.

This will inadvertently and invariably lead to the tremendous evolution of Autonomous Cyber Defense Systems that will act as a counter measure for Antagonist/Adversarial AI powered Powerful Ddos attacks and Polymorphic Malware obfuscation , and also compelling , transubstantiating phishing scams.

It may very well one day protect Advanced Modular Portable Battlefield Management Systems against APT based Cyber Electronic Warfare .

That day I was reading in some military defense journal that Signals Intelligence has also become a Cyber activity. Cyber being the only domain that interacts with Air , land , Space and Maritime operations.

Some tremendous Earth shaking innovations in Military Cyber Space Defense Intelligence Industrial Complex can be observed that analysts can clearly attribute to the last decade or so.

Now if we task Machine based generative Intelligence to infuse zombie nouns in this lucid conversation and make it needlessly dense (from a human readers perspective at least) we may ambiguously conclude that Generative AI is locomoting on the boundaries of self awareness.

Lastly , I beleive that we desperately need Web based Content management platforms (looking at the AI revolution in this Domain) that natively convert rich pdf and presentation documents into web-based HTML for being published on websites and Blogs. WordPress / Drupal(Whitehouse)/ Substack / Medium(NATO) etc.

AI Automation in web development is so essential especially when you convert pdf / presentation docs to Web formats / html. The present state only allows Document embeds. Native document conversion to html must support clickable hyperlinks and covert video links as playable video embeds. Just upload your document/presentation to content management platform like WordPress / medium / Drupal and it will populate the content and resources appropriately in a predefined format..page wise and post wise etc.

Human Intelligence psychic second sighted abilities get influenced all the time by elementary sub atomic particles and EM energy waves that are not a part of our DNA at protoplasm or cytoplasm level whatsoever .

So the credit never all in all go to human alone. Our consciousness and Cognitive abilities are not our alone. They get influenced all the time by imperceptible external factors and forces even in isolation and transcendental states. As not much is known about the lions share of energy and matter observed through the prism and tesseract of elementary subatomic particles that most likely operates in the circumpolar macrocosm.

Therefore the philosophy that 1 can achieve infinite heights if 1 does not concern oneself with who gets the credit for accomplishment of work related project . Acquiring more organisation specific / oriented approach.

*Militarisation / Weaponisation of AI.
United Nations University(unu.edu)
Cyberdefensereview.army.mil
*Contextualizing Deepfakes (Defense.gov)
*Dawn of Cyber Kinetic
Militarisation of CyberSpace

(ccdcoe.org)
*Nuclear Sector Cybersecurity framework
Critical Infrastructure (CISA)
*Autonomous Cyber Defense
(AI & Immune System Approach)
*Threats in Space - UCIGCC.org
UNOOSA.org
*DefenseIntelligenceAgency.mil
*Electromagnetic Spectrum Conflict
(Crows.org)
*Cold War 2 Cyber War - 25 yrs.Springer.
*Computer Security in Nuclear World
IAEA Vol (64-2).
*AI Security Center - National Security Agency.
Army.mil
*Precognition Parapsychology ESP
Clairvoyance - Britannica.com